WP 4 - Graph parallelism & graph types

FeatureCloud

Privacy preserving federated machine learning and blockchaining for reduced cyber risks in a world of distributed healthcare

H2020 - 826078

General goals & objectives

1. Explainable & privacy-aware Machine Learning
2. ... on high-dimensional data
3. ... by using graphs as intuitive & universal data structures to depict topology and relations between entities

Graph-based / hybrid (= in addition to NNs) approaches have several advantages:

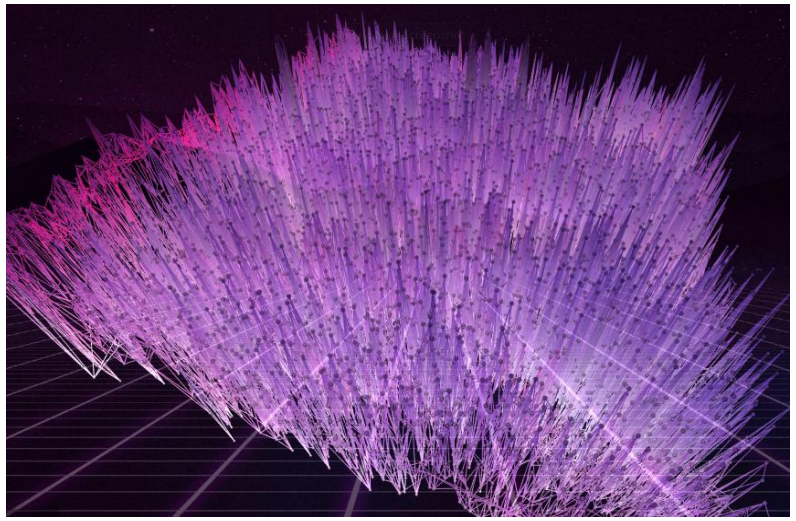
1. Form naturally on many interesting data sets (see next slide)
2. Easier to interpret (in structure as well as results) - in part because they inherently project high-dimensional data into a 2D/3D space
3. Allows for a multi-stage & distributed system where each signal conveys meaning in itself (interest update, recommendation, chemical reaction, ...)

Graphs as universal data structures

Social network

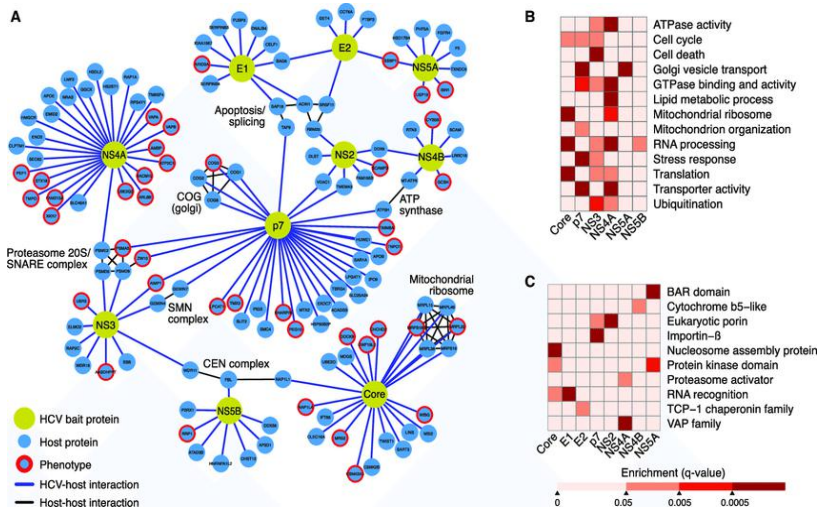
http://socialengineindia.com/

Cell image => graph



Holzinger, Andreas & Malle, Bernd & Giuliani, Nicola. (2014). On Graph Extraction from Image Data. 552-563. 10.1007/978-3-319-09891-3_50.

Protein-protein IN



Ramage, Holly & Kumar, Gagandeep & Verschueren, Erik & Johnson, Jeffrey & Dollen, John & Johnson, Tasha & Newton, Billy & Shah, Priya & Horner, Julie & Krogan, Nevan & Ott, Melanie. (2015). A Combined Proteomics/Genomics Approach Links Hepatitis C Virus Infection with Nonsense-Mediated mRNA Decay. Molecular cell. 57. 329-340. 10.1016/j.molcel.2014.12.028.

Learning on distributed graphs - Challenges

1. **Expensive communication**, since all nodes need to update a central (infrastructure of) servers according to their local model evolution

2. **Systems Heterogeneity**, meaning that edge devices might be of severely different storage & computational capacity

3. **Statistical Heterogeneity**, since each local agent might have their own objective function and subsequently differently distributed data sets (I.I.D. assumption does not hold).

4. **Privacy concerns**, meaning that model updates (even without transmitting the underlying data) can reveal sensitive information, while current approaches to counteract this phenomenon (e.g. adding noise via a differential privacy model) can significantly reduce systemefficiency.

Graph convolutional networks & feature propagation

- Graphs share *“hierarchical feature layers”* with images, but not their spatial locality (rigid grid-pattern)
- In recent years, several methods have been established to learn on huge graphs by propagating & coalescing features amongst nodes
- Although these methods have been successful in large enterprise settings, the question is how to implement them in a distributed way using resources on the “edge” (as opposed to the backbone)
- Also, the theory behind GCN’s is not yet fully established.

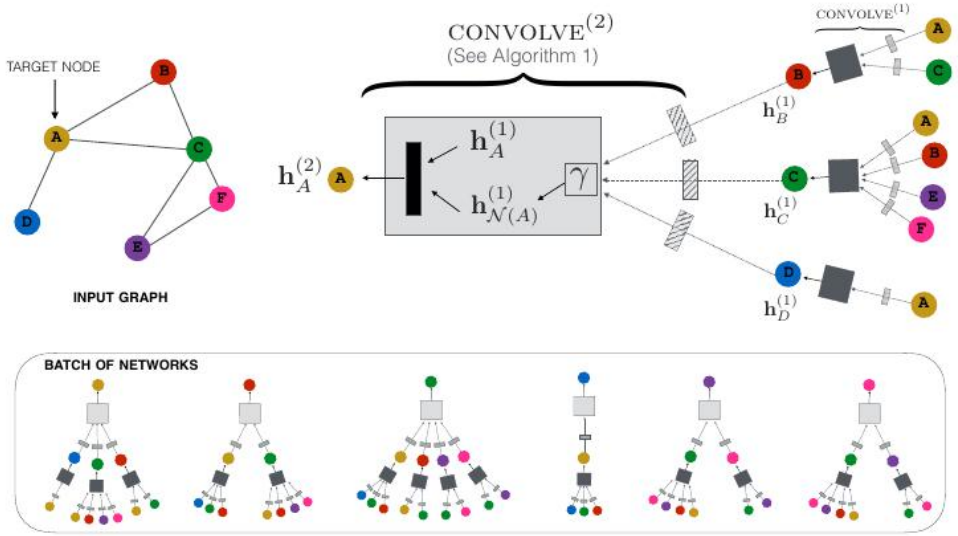


Figure 1: Overview of our model architecture using depth-2 convolutions (best viewed in color). Left: A small example input graph. Right: The 2-layer neural network that computes the embedding $h_A^{(2)}$ of node A using the previous-layer representation, $h_A^{(1)}$, of node A and that of its neighborhood $N(A)$ (nodes B, C, D). (However, the notion of neighborhood is general and not all neighbors need to be included (Section 3.2).) Bottom: The neural networks that compute embeddings of each node of the input graph. While neural networks differ from node to node they all share the same set of parameters (*i.e.*, the parameters of the CONVOLVE$^{(1)}$ and CONVOLVE$^{(2)}$ functions; Algorithm 1). Boxes with the same shading patterns share parameters; γ denotes an importance pooling function; and thin rectangular boxes denote densely-connected multi-layer neural networks.

Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 974–983, 2018. doi: 10.1145/3219819.3219890

Federated Learning - traditional

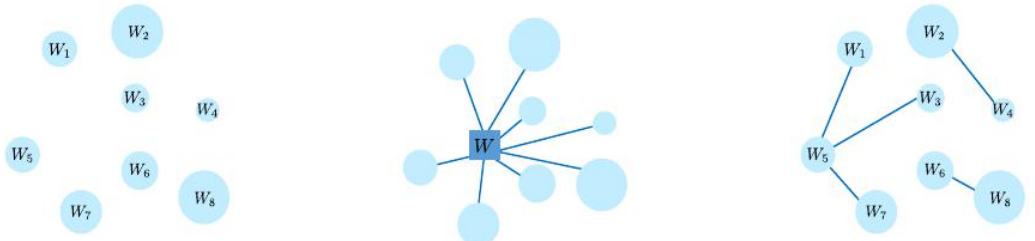
Top image: Traditionally (as introduced by Google), the goal of FL is to learn a global model from distributed data:

1. A global (pre-trained model) is distributed to all client devices -> each one gets exactly the same!
2. As users interact with the client, the model gets updated individually, resulting in a specialized model over time.
3. In intervals, clients compute diffs & send them to the server, the sum of which are reconciled into a new global model, which again is distributed downstream.

Bottom image: 3 possible modes of FL:

1) individual, 2) global, 3) learn from peers

https://blog.ml.cmu.edu/2019/11/12/federated-learning-challenges-methods-and-future-directions/



(a) Learn personalized models for each device; do not learn from peers.

(b) Learn a global model; learn from peers.

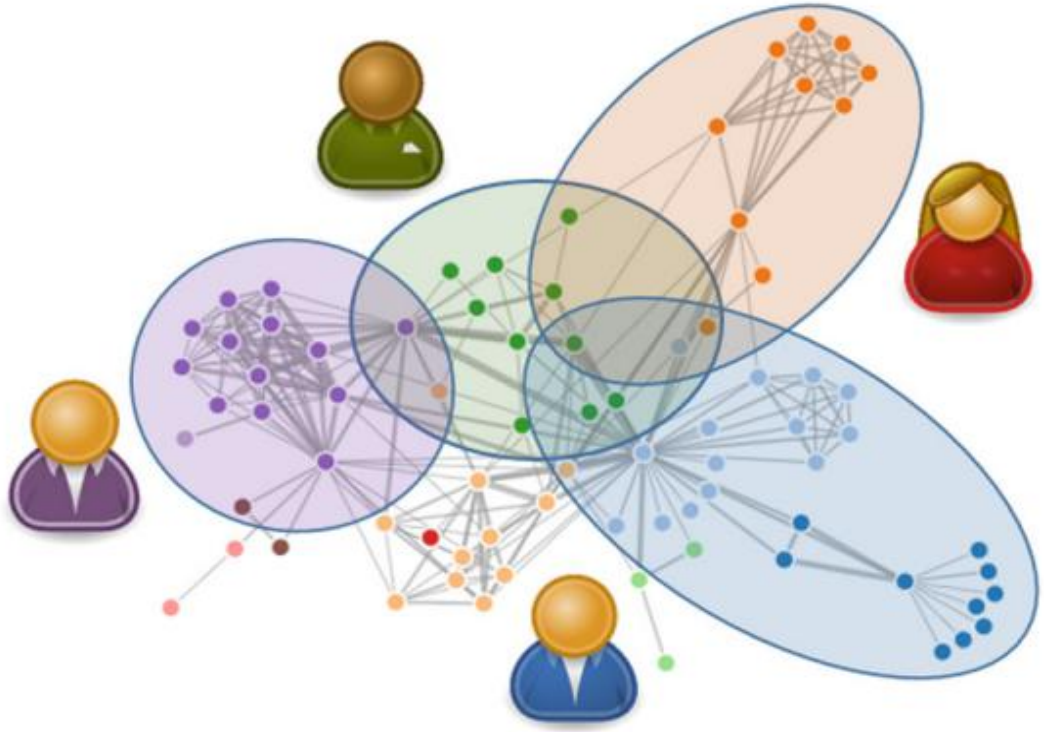
(c) Learn personalized models for each device; learn from peers.

Figure 5: Different modeling approaches in federated networks. Depending on properties of the data, network, and application of interest, one may choose to (a) learn separate models for each device, (b) fit a single global model to all devices, or (c) learn related but distinct models in the network.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated Learning: Challenges, Methods, and Future Directions. pages 1–21, 2019. URL http://arxiv.org/abs/1908.07873.

Local spheres - going beyond federations

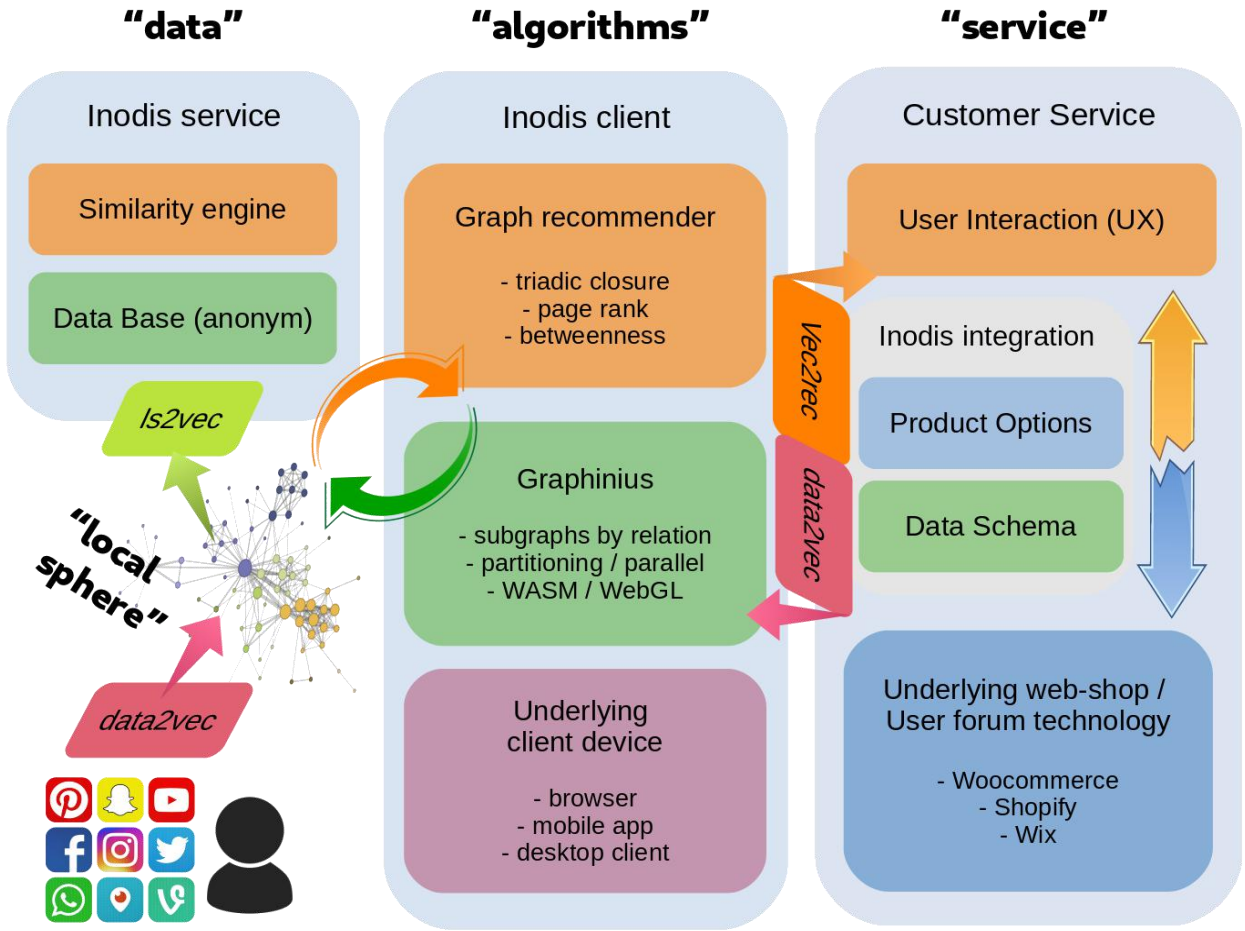
- In human society, solutions to complex problems are usually solved by a collaboration of experts, each contributing their unique talents - NOT by smoothing out the knowledge base of the collective.
- Likewise, allowing each node to retain a local model fitting their respective data / objective function might lead to better global results even in the absence of a global model.
- This way, each local model would act as a re-usable component - and the swarm could adapt to new problems without continually re-training from the start.



Bernd Malle, Nicola Giuliani, Peter Kieseberg, and Andreas Holzinger. The More the Merrier - Federated Learning from Local Sphere Recommendations. In Machine Learning and Knowledge Extraction, IFIP CD-MAKE, Lecture Notes in Computer Science LNCS 10410, pages 367–374. Springer, Cham, 2017. doi: 10.1007/978-3-319-66808-6 24.

Local spheres - local / global connection

- Efficient communication among relevant local spheres is crucial
- We want to avoid “gossip-style” networks (everyone talking to everyone else)
- We propose computing a “fingerprint” per local sphere in the form of a representative feature vector (ls2vec)
- This way, a central similarity service can “connect” local spheres most likely to profit from each other’s local model / expertise.



Bernd Malle - Inodis system architecture

Thank you !

Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554 “A reference model of explainable Artificial Intelligence for the Medical Domain”.

Parts of this work have been funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 826078 , "Feature Cloud".